Researchers keep rejecting grandmother cells after running the wrong experiments: the issue is how familiar stimuli are identified.

Jeffrey S. Bowers

Nicolas D. Martin

Ella M. Gale

School of Psychological Science, University of Bristol, Bristol, BS8 1TU, U.K.

Corresponding Author: Jeffrey S. Bowers, School of Psychological Science, University of Bristol, Bristol, BS8 1TU. Email: j.bowers@bristol.ac.uk

Abstract:

There is widespread agreement in neuroscience and psychology that the visual system identifies objects and faces based on a pattern of activation over many neurons, each neuron being involved in representing many different categories. The hypothesis that the visual system includes finely-tuned neurons for specific objects or faces for the sake of identification – so-called 'grandmother cells' – is widely rejected. Here we argue that the rejection of grandmother cells is premature. Grandmother cells constitute a hypothesis of how familiar visual categories are identified, but the primary evidence against this hypothesis comes from studies that have failed to observe neurons that selectively respond to unfamiliar stimuli. We review these findings and show why they are irrelevant. Neuroscientists need to better understand existing models of face and object identification that include grandmother cells, and then compare the selectivity of these units with single neurons responding to stimuli that can be identified.

1. Introduction

The hypothesis that single neurons mediate the identification of familiar visual categories (e.g., a familiar person) is often dismissed in neuroscience. Indeed, researchers tend to use the somewhat pejorative phrase 'grandmother cell' to describe this form of neural coding. Instead, it is widely claimed that visual categories are identified on the basis of a pattern of activation over many neurons, each neuron being involved in representing many different categories; so-called 'distributed' or 'ensemble' coding. The distributed approach is thought to be supported by theoretical considerations and empirical findings. Indeed, single-cell recording studies are often taken to falsify the grandmother cell hypothesis.[1]

In this article, we detail why the widespread dismissal of grandmother cells is misguided. The core problem is that researchers tend to consider straw-man versions of the hypothesis, and as a consequence, do not carry out experiments relevant to testing more plausible versions of the theory. We focus on one of the most common theoretical confusions that has led to a consistent failure to carry out relevant experiment, namely, the claim that grandmother cells constitute a theory of how both familiar and unfamiliar categories are perceived. Researchers then reject grandmother cells when they do not find neurons that respond selectively to unfamiliar stimuli (e.g., failing to find a neuron that selectively responds to an unfamiliar face). The problem with this, however, is that grandmother cells constitute a theory of how familiar categories are identified (e.g., how you identify your grandmother and how I recognize mine). Accordingly, the relevant question is how neurons respond to familiar faces, objects, and words when they are identified. As we detail below, relatively few studies assess the selectivity of neurons under these conditions, but there is evidence that neural selectivity is greater for familiar than unfamiliar categories. Indeed, given the high-level of selectivity that has been reported with familiar visual categories when they are identified, the grandmother cell hypothesis should be taken as a serious hypothesis rather than dismissed out of hand.

This article is organized as follows. First, we describe in some detail what we mean by the term grandmother cell, and then briefly describe a number of computational models of face, word, and object identification in psychology and neuroscience that implement grandmother cells. Second, we briefly review studies that assessed the response selectivity of single neurons in monkey visual cortex to unfamiliar stimuli (e.g., unfamiliar faces or objects), or familiar stimuli that the monkey was never trained to identify (novel stimuli that the monkey repeatedly viewed without any need to categorize). Some of these studies are reviewed in the text, others are just noted in Table 1. Although these studies rarely report neurons that responded selectively to one stimulus category

(e.g., a specific face), these studies are irrelevant to the grandmother cell hypothesis because grandmother cells constitute a theory of how familiar visual images are identified as members of specific categories. Third, we review several studies that do report highly selective responses in visual neurons when monkeys were tested on familiar stimuli they could identify – exactly the conditions in which grandmother cell theories predict highly selective responding. Finally, we briefly mention computational studies that show there are computational advantages with grandmother cell representations. Together, these empirical and computational studies show that it is premature to reject grandmother cells.

1.1. What is the grandmother cell hypothesis?

A grandmother cell is a neuron that represents a single familiar visual category, either a basic level category (e.g., a cell that represents the visual category 'dog', or 'bicycle') or a subordinate level category (e.g., a cell that represents the visual category 'my bicycle', or a specific person such as 'my grandmother'). On this hypothesis, an object is identified at a basic level or subordinate level when the corresponding cell(s) fires beyond some threshold.[2] Grandmother cells are often compared to highly-sparse representations in which a visual category is represented by a small number of neurons, each neuron only contributing to the representation a few different visual categories. However, the grandmother cell hypothesis is at the extreme end of the sparseness continuum: a single neuron represents a single visual category.[3] It should be noted that grandmother cell theories are not committed to the claim that all neurons are grandmother cells, and indeed, some combination of grandmother cells and distributed representations may work hand-in-hand in some tasks. By contrast, critics of grandmother cell theories reject the hypothesis that there are any grandmother cells. Within psychology, the equivalent to a grandmother cell is a localist representation in which single units in models encode specific visual categories at a basic or subordinate level. In this situation, the model identifies inputs by activating single units above some threshold. Unlike grandmother cells in neuroscience, localist representations in psychology are not couched in neuroscience terms, but functionally, grandmother cells and localist representations are equivalent.[4-6] Grandmother cells and localist representations constitute a hypothesis about how all basic and subordinate visual categories are identified, not only about how we identify faces at a subordinate level.

1.2. Understanding Grandmother Cells Through Modelling.

In order to better understand the grandmother cell hypothesis, it is worth briefly reviewing a few models in computer science, neuroscience and psychology that include grandmother cells (or their equivalent localist coding) in order to identify familiar visual categories at a basic or subordinate level.

The initial motivation for many of these models was the observation that the visual system is hierarchical: 'simple' cells in the visual cortex code for line orientations at a given retinal location, and neurons in subsequent layers of the hierarchy code for more abstract (e.g., more spatially invariant) and complex stimuli.[3] According to the grandmother cell hypothesis, this process of abstraction continues over multiple layers, the neurons at the top of the visual hierarchy coding for complete objects or persons in such a way that a single neuron responds more strongly to images of this category compared to others. Hubel himself considered whether grandmother cells are the natural extension of his theory of early vision, writing:

> "What happens beyond the primary visual area, and how is the information on orientation exploited at later stages? Is one to imagine ultimately finding a cell that responds specifically to some very particular item (Usually one's grandmother is selected as the particular item, for reasons that escape us.) Our answer is that we doubt there is such a cell, but we have no good alternative to offer." (p. 96)[7]

1.3. Grandmother cells in computer science and neuroscience models

Although Hubel doubts that there are cells that code for visual categories at the top of the visual hierarchy, it is important to note that there are many computational models that have just this property. An early example in computer science was the "neocognitron" model[8] that was directly inspired by Hubel and Wiesel's model[3] of simple and complex cells in primary visual cortex (V1). The neocognitron model roughly simulated the early processes in V1 and continued the hierarchy of processing steps to a top layer that coded each familiar category (written numbers between 0-9) with a single unit that selectively respond to one number. Indeed, the authors used the term 'grandmother cell' to describe these selective units in their model. The neocognitron was in turn one of the inspirations for computational models in neuroscience that include units that selectively represent familiar categories, such as a face of a specific person (e.g., [9]).

More recently, Thorpe and colleagues[10, 11] have been developing hierarchical artificial neural networks that incorporate the biologically plausible learning rule spike-time-dependent-plasticity (STDP). STDP is an unsupervised form of learning that adjusts the connection strengths based on the relative timing of a particular neuron's output and input action potentials. Importantly for present purposes, models of object identification that employ STDP often learn highly selective units, as illustrated in Figure 1a. In multiple papers, Thorpe highlights the computational advantages of highly selective representations and explicitly relates these selective representations to grandmother cells. For example, Thorpe wrote:

> "I will discuss how a combination of STDP and temporal coding can allow highly selective responses to develop to frequently encountered stimuli. Finally, I will argue that 'grandmother cell' coding has some specific advantages not shared by conventional distributed codes."[13]

1.4. Localist models in psychology

Within psychology, the Interactive Activation (IA) model of visual word identification is an example of an early and influential theory that coded for familiar categories (words) with single units.[14] The IA model was again hierarchically organized with localist letter features at the input layer, localist letter detectors at a second layer, and localist word detectors at the output layer. In this scenario, a specific word is recognized when its localist word unit is activated beyond some threshold. Localist representations are commonplace in theories in psychology, including in models of visual word identification,[15] models of spoken word identification,[16] spoken word production,[17] face perception,[18] and many other domains of theorizing.

Despite the existence of the above models, it is important to emphasize that most theorists in psychology and neuroscience strongly reject grandmother cells. For example, within psychology, the localist word representations in the IA model are often rejected in favour of distributed representations within the Parallel Distributed Processing (PDP) models of word identifications (e.g., [19]). Even researchers who endorse localist codes in psychological models are rarely committed to the claim that their model is implemented with grandmother cells – instead, the models are considered theories of cognition with no commitments to neurophysiology. Furthermore, in computational neuroscience where the models are intended to make claims about neural coding, it is very much a minority position to endorse grandmother cells. The far more common claim is that single neurons do not code for specific categories.[20, 21] As Averbeck, Latham, and Pouget put it:

> "As in any good democracy, individual neurons count for little; it is the population of activity that matters". (p.358).[22]

Similarly, researchers who carry out single-cell recording studies overwhelmingly reject grandmother cells. But as we show below, this conclusion is compromised by the common assumption that grandmother cells should code for unfamiliar visual categories.

To summarize then, a grandmother cell theory is committed to the claim that a subset of neurons selectively represents familiar visual categories at the basic level and at the subordinate level, and that different grandmother cells encode basic and subordinate visual categories (one cell coding for the basic visual category 'bicycle' and another cell coding for the subordinate 'my bicycle'). These selective representations are hypothesized to support the

identification of these familiar stimuli when their activation goes beyond some threshold. The grandmother cell hypothesis does not reject distributed representations (both forms of representations may play a role in vision), whereas critics of the grandmother cell hypothesis claim that grandmother cells do not exist.

1.5. The theoretical confusion that leads to the premature rejection of grandmother cell theory.

Bowers has highlighted a number of confusions regarding grandmother cells that have led researchers to reject only straw-man versions of the hypotheses,[5] such as the claim that there is one grandmother neuron per visual category (such that a loss of one neuron can lead to a failure to recognize your grandmother),[23] and the claim that grandmother cells respond to one category of input and remain entirely silent to all other categories (contrary to the common observation that a neuron responds above baseline to images from multiple categories.[24, 25] But there is no reason to adopt these characterizations of grandmother cell theory. Indeed, the grandmother/localist representations in the various models described above do not have the property that they only respond to one category of input and are silent to everything else. Rather they have the property that they respond most strongly to one category, and less strongly to visually similar inputs from different categories. Of course, visually similar inputs should not activate a grandmother/localist unit beyond the threshold for identification, and if they do, it would result in a misidentification of the input. For more detailed discussion and debate on this issue see ([5, 26, 27, 28]). Similarly, multiple redundant grandmother cells are consistent with grandmother cell theory,[4, 5, 25] and indeed, redundant localist codes can emerge in artificial neural network models after training,[29] making the model more robust to the removal of a single unit.

But here we focus on the false claim that grandmother cell theories should explain how we perceive and identify unfamiliar things. This leads researchers to falsely reject grandmother cells on the basis that there are just not enough neurons in the brain to code for all possible unfamiliar stimuli,[27, 30] and most critical for present purposes, leads researchers to falsely reject grandmother cells on the basis of not finding neurons that respond selectivity to unfamiliar stimuli.

2. What does the research show?

In order to review as many studies as possible we describe some of the most relevant papers in the text and list other key studies in Table 1. In many or most studies reviewed below the monkeys were repeatedly presented with the stimuli used to drive visual neurons, either before the recording session or during the recording session itself. For this reason, the authors of these studies

often report that the stimuli tested were familiar. However, in most of these studies the monkeys simply saw the images passively without making any response, or they performed some task that did not require them to identify the images at the basic or subordinate level. Accordingly, there no reason to assume that the monkeys were able to identify the images at subordinate or basic levels. Our main point is that if animals cannot identify the stimuli, there is no reason reject grandmother cells when single neurons fail to respond selectively to these stimuli.

2.1. Chang and Tsao (2017)

We start with the Chang and Tsao study[31] because it has received a lot of attention, makes strong claims in support of distributed over grandmother cell coding for face identification, and it is the most recent of a series of papers by Tsao and colleagues on the neural coding of faces (see Table 1 for other related studies from this group).

The authors showed two monkeys images of faces while recording from neurons from the middle lateral (ML)/middle fundus (MF) and anterior medial (AM) regions of Inferotemporal (IT) cortex that have previously been shown to be involved in different aspects of face processing. The faces were synthetic human faces that had been transformed in high-dimensional space in order to assess the impact of various transformations on neural firing (See Figure 2a for example faces). The faces were presented for 150ms, interleaved with 150ms of a grey screen, each image being presented between three and five times. The stimuli included 2000 images of parametrized frontal faces, 2000 images of parameterized profile faces, and the task of the monkeys was simply to fixate on a fixation point for a juice reward, and accordingly, the faces were just viewed passively. They reported that a pattern of activation over approximately 200 neurons coded for specific faces and took the findings to support distributed rather than grandmother cell coding. In the "In Brief" summary of their article they wrote:

> "Facial identity is encoded via a remarkably simple neural code that relies on the ability of neurons to distinguish facial features along specific axes in face space, **disavowing the long-standing assumption that single face cells encode individual faces.**" [31] [bold added]

But no theory (grandmother or otherwise) should predicts that single cells in monkey brains should selectively respond to unfamiliar human faces. In the same way, no theory should expect to find neurons in human visual cortex that selectively responds to an unfamiliar monkey face.

The conceptual confusion of the authors was highlighted more strongly in the interview with Doris Tsao who said the following in an interview linked with the Chang and Tsao (2017) paper:

> "Before this work that is described in this paper itself... people thought at the highest levels of the brain's face recognition system there are cells that are selective for specific individuals, all the people that you know and recognize there are cells encoding them.
>
> And obviously this raised a question, which is how can one have enough cells to represent all the people that you possible could recognize. **There are 6 billion people on this earth, and obviously you do not have 6 billion cells specialized for face recognition in your brain**. So it was a mystery how it is ultimately done." [32] [bold added]

The conclusion Tsao draws is that the brain must rely on distributed coding rather than grandmother cells. But given this conception of grandmother cells, it raises the question why the authors even bothered using their data to argue against grandmother cells given that it was ruled out *a priori* on the basis that there are not enough specialized neurons in a monkey brain. Nevertheless, this same conclusion was drawn by Quian Quiroga in a "Leading Edge Previews" article in the same issue entitled "How Do We Recognize a Face?"[1]

He writes:

> "As the authors argue, their results imply that there are no detectors for face identity at the single neuron level in the face patch system and, consequently, **this may put an end to the long-standing dispute about the existence of grandmother cells in visual cortex**."[1] [bold added]

2.2. Rolls (2017)

Rolls, in a special issue of a journal devoted to the topic of grandmother cells, reviewed a number of findings from his lab that he takes to be falsify the grandmother cells.[21] As he writes in the abstract: "The encoding of information in the primate inferior temporal visual cortex, hippocampus, orbitofrontal cortex, and insula is described. All these areas have sparse distributed graded firing rate representations". However, once again, this conclusion come from studies in which animals were presented with unfamiliar stimuli, or where the familiarity of the stimuli where not discussed, and where the animal did not need to distinguish images one from another.

For example, the first study he summarizes was carried out by Rolls and Tovee that assessed the neural coding of face and object identity in the temporal visual cortex.[33] This highly cited paper assessed the firing selectivity of 14 neurons to a set of 68 stimuli. None of the neurons responded selectively to one of the stimuli, and Rolls takes the findings as inconsistent with the grandmother cell hypothesis.[21] This is a strong conclusion to reach on the basis of recording from 19 neurons, but more relevant for present purposes, the stimuli were composed of unfamiliar human and monkey faces and unfamiliar non-face

stimuli that included images of woodland, countryside, and foods: some of these images are reproduced in Figure 2b. The monkey task was simply to look at the photos without responding (the monkeys received reward by licking a tube between image trials), and accordingly, there was no pressure to learn to the images. Given that the monkey was not trained to identify the stimuli, the experiment does not test the grandmother cell hypothesis.

A search of Google Scholar using search terms 'grandmother' and 'Rolls' as author identifies 87 papers, and as far as we can tell there is only one paper where there was a brief note that selectivity might be greater for familiar categories,[34] but this was limited a single paragraph where they simply note that their findings "raise the possibility" (p. 213) that visual experience my impact on the tuning of single neurons.

2.3. Lehky Kiani, Esteky, and Tanaka (2011)

Lehky *et al*. recorded from 674 monkey inferotemporal cells, each stimulated by 806 object photographs.[35] Although they found some neurons to respond highly selectively, they failed to obtain evidence for grandmother cells, writing:

> "We believe that the data presented here do not support 'grandmother cell' coding.... On average, the second largest response was almost the same size as the largest response (89.3% for single-neuron responses, 79.5% for population responses). This is not a characteristic of 'grandmother cell' coding."[35]

But again, the conclusion is not justified given that the stimuli were not familiar and the monkeys were not trained to discriminate between the stimuli (see Figure 2c for some example images from this experiment). Interestingly, the authors do leave open to door for grandmother cells for familiar, meaningful stimuli, writing: "Nevertheless, it is still possible that grandmother cell encoding could occur for a small number of special objects to which the observer was highly exposed, and which also had strong behavioral significance" (p. 1115). Why such selective coding should be restricted to a few categories with strong behavioural significance as opposed to all visual categories that can be identified is unclear.

2.4. Rust and DiCarlo (2012)

Rust and DiCarlo designed a study to measure V4 and IT neuronal sparseness using a set of 300 natural images that included an object in its natural context, each object being distinct (it would be called by a different name).[36] The authors wrote that: "Images included a wide variety of content, including objects familiar to the animal, other (unfamiliar) animals, man-made objects, other monkeys, and people", but provided no analyses of whether the results

were impacted on familiarity. But the stimuli reported in their figures suggest that the vast majority of images would be unfamiliar, and a few example images are depicted here in Figure 2d. Indeed, the authors appear to consider familiarity of the images a problem, writing: "To guard against possible non-stationary effects (e.g., familiarity with the images), recordings were altered between V4 and IT" (p., 10173). Again, the monkeys were not required to discriminate between any of these images. Based on the fact that they did not find any neurons that selectively responded to the test images, the authors argue that the visual system codes information in a distributed manner.

But again, no grandmother cell theory would predict selective responding to the unfamiliar images, and the authors did not specifically discuss or analyse the results for the familiar images in the experiment. As summarized in Table 1, other studies have also failed to observe highly selective responding of IT neurons in response to stimuli that a monkey was not trained to identify.

There is yet another issue with some of the above studies that further weakens the conclusions that have been drawn regarding grandmother cells. That is, there is growing evidence that monkeys are quite poor at recognizing faces, and that they recognize faces in a qualitatively different way than humans.[37] The findings include the observation that monkeys require extensive exposure and training in order to reach only moderate performance in identifying specific faces in a laboratory setting, performance on matching tasks is worse for faces than for other objects, and monkeys are no better at matching familiar compared to unfamiliar faces (unlike humans). In addition, Macaques show no inversion effects, have no distinct ventral face-specific pathway, and no right hemisphere specialization for faces. This further undermines the common rejection of grandmother cells based on studies that did not even train monkeys to identify (unfamiliar) faces.

3. Highly selective representations are more often found when familiar stimuli are tested.

A number of studies[38-41] have reported that the selectivity of neural firing increases as the stimuli become more familiar to the animal and this alone suggests that the above results provide a poor basis for characterizing the representations that support object and face identification. But the important question for present purposes is whether stimuli that can be identified at a basic or subordinate level evoke levels of selectivity in visual neurons that are consistent with grandmother cells.

The most high-profile set of experiments that have reported highly selective neural firing in response to familiar stimuli were recorded from the hippocampus and related structures in humans (e.g., the 'Jennifer Aniston' neurons[42]). It is important to note, however, that these neurons are part of the

memory rather than the visual system, and accordingly, the results are not directly relevant to the classic grandmother cell hypothesis that is concerned with vision. Nevertheless, it raises the question is whether similar levels of selectivity can be obtained with familiar stimuli within the visual system.

In fact, as reviewed below, a few studies have reported highly selective responding in IT cells when the monkey is presented with stimuli they can identify, and in one case, the level of selectivity is as high as any selectivity observed in the hippocampus. Although many of studies are quite old, they are highly relevant today. Unfortunately, the central message that selectivity can be extremely high for familiar objects in visual cortex seems to have been lost on the many recent experimental single-cell recording studies reviewed above as well as recent theoretical discussions of these findings,[1, 43, 44] but see [45] for a counter-example.

3.1. Logothetis, Pauls, and Poggio (1995)

Logothetis, Pauls, and Poggio trained two monkeys to identify over 100 novel computer-generated objects from various viewpoints over the course of months.[46] After learning a substantial subset of stimuli, the monkeys performed a visual matching task in which they first fixated at a target stimulus from one viewpoint and then saw a series of test stimuli from various viewpoints that were from the same class or not, as illustrated in Figure 2e. The monkeys categorized the objects as matching or mismatching while the authors recorded from 796 neurons in the upper bank of the anterior medial temporal sulcus. The neurons showed a range of selectivity, a few (3/796; 0.37 %) responding selectively to only one object presented from any viewpoint, and a larger set (93/796; 11.6 %) responding selectively to a subset of views of one of the known target objects but less frequently (or not at all) to highly similar objects. These are perhaps the most selective responses ever recorded in any part of the brain. Critically, no selective responses were encountered for views that the animal systematically failed to recognize. That is, highly selective codes were associated with the ability to identify the object. Furthermore, the authors found that the percentages of cells responding to objects from a given class correlated with the amount of training. This led the authors to write: “Thus, it seems that neurons in this area may develop complex configurational selectivity as the animal is trained to recognize specific objects”.[46]

3.2. Sakai, Naya and Miyashita (1994)

Sakai *et al*. trained monkeys to recognize 12 pairs of computer-generated Fourier patterns while recording from 474 neurons in Anterior Inferior Temporal cortex.[47] One of the two stimuli from a paper (the “cue”) was presented for .5 seconds and after a five second delay the paired associate image or a foil image were presented side-by-side and monkey had to select the

associate pattern to receive an award. This required monkeys to learn and discriminate these patterns. The authors identified 89 neurons that responded most strongly to one of the cue stimuli. The Fourier pattern cue stimuli were then systematically manipulated in order to assess the impact of varying the learned images on neural responses. In the vast majority of cases, the neurons responded more strongly to the trained pattern compared to the untrained transformed ones, and in no case did the neuron respond more strongly to the transformed pattern. This suggests that the neurons were tuned to the trained visual patterns. This is exactly as one might expect from a grandmother cell theory. See Figure 2f for example of trained images and manipulated images.

3.3. Perrett, Smith, Potter, Mistlin, Head, Milner and Jeeves (1985)

Perrett e*t al*. recorded from cells in the superior temporal sulcus of macaque monkey in studies that were designed to distinguish between faces from non-face stimuli (that is the general category of face rather than the identity of a specific face).[48] But during testing the authors noted that some cells that responded to one familiar person more than others. For instance, one cell responded more to a wide range of views of one familiar person (Paul Smith, one of the experimenters), compared to another (David Perrett, another experimenter), such that the identity of the person (based on whole body, not just face) could be determined by the activation of this one neuron. Perrett and colleagues took these findings to highlight the role of learning on neural selectivity, and the that their findings were at least consistent with grandmother cell coding schemes, writing:

> "The responses of these cells have many of the properties hypothesized for 'gnostic units' and provide insight into the final stages of visual processing leading to the recognition of an object as a face and more specifically the identity of the fact."[49]

Note, the term 'gnostic unit' is being used there the same way as grandmother cell.

3.4. Kobatake, Wang, and Tanaka (1998)

Kobatake *et al*. trained 2 monkeys to discriminate between 28 moderately complex shapes and then recorded from 131 cells in inferotemporal cortex. A single shape (for examples see Figure 2) was displayed on a computer screen and it disappeared when it was touched. After a short delay the stimulus was displayed along with four additional foils from the same set of 28 stimuli. The task was to select the repeated stimulus. Training was extensive and was completed when the monkeys performed 500 successful trials per day with a success rate of over 75%. Neural recordings to these stimuli were performed while these two trained monkeys were under anaesthesia, as well as 130

inferotemporal neurons in two additional monkeys who were not trained to identify and remember the stimuli.

For the trained monkeys there were 28 cells that maximally responded to one of the trained stimuli, and for these neurons, an average of 3 other stimuli from the trained set evoked responses greater than 50% of individual cells' maximal response. Based on this the authors concluded "The broad tuning may suggest that the discrimination depended on activity of cell population". Unfortunately, however, the authors did not report whether any of the 26 critical failed to respond strongly to other stimuli (the overall average of 3 stimuli evoking strong response may well have included some neurons that were more selective). The failure to report the most selective responding neuron is unfortunate. Importantly, the authors also reported the selectivity of the neurons was much greater for the trained compared to non-trained monkeys, highlighting the fact that selectivity is greater for stimuli that can be identified.

3.5. Khuvis, Yeagle, Norman, Grossman, Malach and Mehta (2018)

As far as we are aware, Khuvis *et al*. have provided the first test of whether individual neurons in humans visual cortex code for specific familiar faces.[50] The authors recorded from 63 neurons in ventral temporal cortex across eight patients. A set of images of 10 famous faces as well as 10 body parts, 10 houses, 10 tools, and 10 abstract patterns and were presented centrally and participants performed a 1-back task during recording in which they indicated whether current stimulus matched the previous one. Twenty-six neurons were found to be selective for the general category faces, and 7 for non-face categories. However, within the face category, the authors did not report any neurons that selectively responded to one face. It will be important to see whether similar results are obtained in humans in follow-up studies, especially it is possible to test a wider range of neurons and face images.

We want to be clear that we do not take reports of extremely selectivity in visual neurons in response to trained stimuli as evidence in support of the grandmother cell hypothesis. There are too few relevant studies, the results are mixed, and it perfectly plausible that the most selective neurons are not so selective that the they are tuned to maximally respond to one visual category. Nevertheless, the most striking demonstrations of selectivity are consistent with the grandmother cell hypothesis, and in our view, it is premature to reject grandmother cells based on current findings.

4. Conclusions and Outlook

We hope that most readers will accept our argument that it is inappropriate to reject grandmother cells on the basis of testing stimuli that cannot be identified, and it is an important observation that neural selectivity is increased when familiar stimuli are tested. At the same time, we expect many

will continue to dismiss the grandmother cell hypothesis. The notion that amongst the billions of neurons in a brain there is a subset of neurons tuned to code for specific categories might seem implausible. We understand this intuition, but it is worth emphasizing again that there are existing models of visual object recognition that are designed to be biologically plausible that have grandmother cells.[9] Indeed, there are biologically plausible models of visual object identification that learn grandmother cells.[12] Even artificial neural network models that are claimed to learn distributed codes in fact learn grandmother representations when trained to co-activate multiple words (or objects or faces) at the same in short-term memory.[29, 51] These findings suggest that there are computational advantages of grandmother cells in the context of short-term memory, much like there are computational advantages of learning highly selective representations in the hippocampus for the sake of episodic memory.[52] Grandmother cells have also been observed in artificial neural network models that learn to identify faces.[53]

In our view, these computational considerations, in combination with the empirical findings above, should lead researchers and theorists to take the grandmother cell hypothesis more seriously. Going forward, researchers need a better understanding of existing models of face and object identification that include grandmother cells, the conditions in which artificial neural networks learn grandmother cells,[29, 51] and then design experiments that compare the selectivity of these units to the selectivity of visual neurons responding to stimuli that can be identified.

References

[1] Quiroga RQ. How do we recognize a face? *Cell.* **2017**, *169*, 975.

[2] Bowers J. Grandmother cells and localist representations: a review of current thinking. *Lang. Cogn. Neurosci.* **2017**, *32*, 257.

[3] Hubel DH, Wiesel TN. Receptive fields of single neurones in the cat's striate cortex. *Am. J. Physiol.* **1959**, *148*, 574.

[4] Page M. Connectionist modelling in psychology: A localist manifesto. *Behav. Brain. Sci.* **2000**, *23*, 443.

[5] Bowers JS. On the biological plausibility of grandmother cells: implications for neural network theories in psychology and neuroscience. *Psychol. Rev.* **2009**, *116*, 220.

[6] Bowers JS. Parallel distributed processing theory in the age of deep networks. *Trends. Cogn. Sci.* **2017**, *21*, 950.

[7] Hubel DH, Wiesel TN. Brain mechanisms of vision. *Sci. Am.* **1979**, *241*, 150.

[8] Fukushima K, Miyake S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In: Competition and cooperation in neural nets Springer; **1982**, 267–285.

[9] Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **1999**, *2*, 1019.

[10] Guyonneau R, VanRullen R, Thorpe SJ. Neurons tune to the earliest spikes through STDP. *Neural Comput.* **2005**, *17*, 859.

[11] Masquelier T, Thorpe SJ. Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Comput. Biol.* **2007**, *3*, e31.

[12] Kheradpisheh SR, Ganjtabesh M, Thorpe SJ, Masquelier T. STDP-based spiking deep convolutional neural networks for object recognition. *Neural Networks*, **2018**, *99*, 56.

[13] Thorpe SJ. Grandmother cells and distributed representations. Visual population codes: Toward a common multivariate framework for cell recording and functional imaging **2011**;p. 23–51.

[14] McClelland JL, Rumelhart DE. An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychol. Rev.* **1981**, *88*, 375.

[15] Davis CJ. The spatial coding model of visual word identification. *Psychol. Rev.* **2010**, *117*, 713.

[16] Hannagan T, Magnuson JS, Grainger J. Spoken word recognition without a TRACE. *Front. Psychol.* **2013**, *4*, 563.

[17] Dell GS. A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* **1986**, *93*, 283.

[18] Burton AM, Bruce V, Johnston RA. Understanding face recognition with an interactive activation model. *Brit. J. Psychol.* **1990**, *81*, 361.

[19] Seidenberg MS, McClelland JL. A distributed, developmental model of word recognition and naming. *Psychol. Rev.* **1989**, *96*, 523.

[20] Churchland PS, Sejnowski TJ, Poggio TA. The computational brain. MIT press; *2016*.

[21] Rolls ET. Cortical coding. *Lang. Cogn. Neurosci.* **2017**, *32*, 316.

[22] Averbeck BB, Latham PE, Pouget A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **2006**, *7*, 358.

[23] Lytton WW. From computer to brain: foundations of computational neuroscience. Springer Science & Business Media; **2007**.

[24] Waydo S, Kraskov A, Quiroga RQ, Fried I, Koch C. Sparse representation in the human medial temporal lobe. *J. Neurosci.* **2006**, *26*, 10232.

[25] Gross CG. Genealogy of the "grandmother cell". *Neuroscientist,* **2002**, *8,* 512.

[26] Bowers JS. More on grandmother cells and the biological implausibility of PDP models of cognition: A reply to Plaut and McClelland (2010) and Quian Quiroga and Kreiman (2010). *Psychol. Rev.* **2010**, *117*, 300.

[27] Plaut DC, McClelland JL. Locating object knowledge in the brain: Comment on Bowers's (2009) attempt to revive the grandmother cell hypothesis. 2010; *Psychol. Rev.* **2010**, *117*, 284.

[28] Quian Quiroga R, Kreiman G. Measuring sparseness in the brain: Comment on Bowers (2009). 2010; *Psychol. Rev.* **2010**, *117*, 291.

[29] Bowers JS, Vankov II, Damian MF, Davis CJ. Neural networks learn highly selective representations in order to overcome the superposition catastrophe. *Psychol. Rev.* **2014**, *121*, 248.

[30] Eichenbaum H. Egram In: Winn P, editor. Dictionary of biological psychology; **2003**, 558-559.

[31] Chang L, Tsao DY. The code for facial identity in the primate brain. *Cell* **2017**, *169*, 1013.

[32] Tsao, DY. PaperClip Interview with author Dr. Doris Tsao, *Cell*, **2017**, *169*, Retreived April 29, 2019, from https://www.cell.com/cell/fulltext/S0092-8674(17)30538-X

[33] Rolls ET, Tovee MJ. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J. Neurophysiol.* **1995**, *73*, 713.

[34] Rolls E. Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Hum. Neurobiol.***1984**, *3*, 209.

[35] Lehky SR, Kiani R, Esteky H, Tanaka K. Statistics of visual responses in primate inferotemporal cortex to object stimuli. *J. Neurophysiol.* **2011**, *106*, 1097.

[36] Rust NC, DiCarlo JJ. Balanced increases in selectivity and tolerance produce constant sparseness along the ventral visual stream. *J. Neurosci.* **2012**, *32*, 10170.

[37] Rossion B, Taubert J. What can we learn about human individual face recognition from experimental studies in monkeys?. *Vision Res.* **2018,** *13*.

[38] Erickson CA, Jagadeesh B, Desimone R. Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys. *Nat. Neurosci.* **2000**, *3*, 1143.

[39] Kobatake E, Wang G, Tanaka K. Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J. Neurophysiol.* **1998**, *80*, 324.

[40] Li N, DiCarlo JJ. Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science.* **2008**, *321*, 1502.

[41] Riesenhuber M, Glezer LS. Evidence for rapid localist plasticity in the ventral visual stream: the example of words. *Lang. Cogn. Neurosci.* **2017**, *32*, 286.

[42] Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I. Invariant visual representation by single neurons in the human brain. *Nature*. **2005**, *435*, 1102.

[43] Clark W.J., Colombo M. (**2018**) Face-Selective Neurons: Comparative Perspectives. In: Vonk J., Shackelford T. (eds) Encyclopedia of Animal Cognition and Behavior. Springer, Cham.

[44] Conway BR. The organization and operation of inferior temporal cortex. *Annu. Rev. Vis. Sc.* **2018**, *4*, 402.

[45] Guntupalli JS, Gobbini MI. Reading faces: from features to recognition. *Trends. Cogn. Sci.* **2017**, *21*, 915.

[46] Logothetis NK, Pauls J, Poggio T. Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* **1995**, *5*, 552.

[47] Sakai K, Naya Y, Miyashita Y. Neuronal tuning and associative mechanisms in form representation. *Learn. Memory.* **1994**, *1*, 83.

[48] Perrett DI, Smith P, Potter D, Mistlin A, Head A, Milner AD, & Jeeves, M. A. Visual cells in the temporal cortex sensitive to face view and gaze direction. *P. R. Soc. London*. **1985**, *223*, 293.

[49] Perrett DI, Mistlin AJ, Chitty AJ. Visual neurones responsive to faces. *Trends. Neurosci.* **1987**, *10*, 358.

[50] Khuvis S, Yeagle EM, Norman Y, Grossman S, Malach R, Mehta AD. Face-selective units in human ventral temporal cortex reactivate during free recall. *BioRxiv.* **2018** Jan 1:487686.

[51] Bowers JS, Vankov II, Damian MF, Davis CJ. Why do some neurons in cortex respond to information in a selective manner? Insights from artificial neural networks. *Cognition*. **2016**, *148*, 47.

[52] Marr D, Willshaw D, McNaughton B. Simple memory: a theory for archicortex. In: From the Retina to the Neocortex. Springer; **1991**.p. 59–128.

[53] Vankov II, Bowers JS. Do arbitrary input–output mappings in parallel distributed processing networks require localist coding?. *Lang. Cogn. Neurosci.* **2017**, *32*, 392.

[54] Dubois, J., de Berker, AO., & Tsao, DY. Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J. Neurosci*. **2015**, *35*, 2791.

[55] Meyers, EM., Borzello, M., Freiwald, WA., & Tsao, D. Intelligent information loss: the coding of facial identity, head pose, and non-face information in the macaque face patch system. *J. Neurosci*. **2015**, *35*, 7069.

[56] Taubert, J., Van Belle, G., Vanduffel, W., Rossion, B., & Vogels, R. Neural correlate of the thatcher face illusion in a monkey face-selective patch. *J. Neurosci*. **2015,** *35*, 9872.

[57] Issa, EB, & DiCarlo, JJ. Precedence of the eye region in neural processing of faces. *J. Neurosci*. **2012**, *32*, 16666.

[58] Freiwald, WA, & Tsao, DY. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, **2010,** *330*, 845.

[59] Meyers, EM, Freedman, DJ, Kreiman, G., Miller, EK, & Poggio, T. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J. Neurophysiol.*, **2008**, *100*, 1407.

[60] Franco, L., Rolls, ET., Aggelopoulos, NC, & Jerez, JM. Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex. *Biol. Cybern.*, **2007,** *96*, 547.

[61] Tsao, DY, Freiwald, WA, Tootell, RB, & Livingstone, MS. A cortical region consisting entirely of face-selective cells. *Science*, **2006**, *311*, 670.

[62] Freedman, DJ, Riesenhuber, M., Poggio, T., & Miller, EK. Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cereb. Cortex.*, **2005,** *16*, 1631.

[63] Keysers, C., Xiao, DK, Földiák, P., & Perrett, DI. The speed of sight. *J. Cognitive Neurosci.*, **2001**, *13*, 90.

[64] Sheinberg, DL & Logothetis, NK. Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J. Neurosci.*, **2001**, *21*, 1340.

[65] Tamura, H., & Tanaka, K. Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cereb. Cortex*, **2001,** *11, 384.*

[66] Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, **1999**, *400*, 869.

[67] Booth, MC & Rolls, ET. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb. Cortex*, **1998**, *8*, 510.

[68] Rolls, ET, Treves, A., & Tovee, MJ. The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Exp. Brain Res.*, **1997**, *114*, 149.

[69] Higuchi, SI, & Miyashita, Y. Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *P. Natl. Acad. Sci*, **1996**, *93*, 739.

[70] Rolls, ET, & Tovee, MJ. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J. Neurophysiol*, **1995**, *73*, 713.

[71] Young, MP, & Yamane, S. Sparse population coding of faces in the inferotemporal cortex. *Science*, **1992**, *256*, 1327.

[72] Desimone, R., Albright, TD, Gross, CG, & Bruce, C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* **1984,** *4*, 2051.

[73] Perrett, DI, Rolls, ET, & Caan, W. Visual neurones responsive to faces in the monkey temporal cortex. *Exp. Brain Res.*, **1982**, *47*, 329.

Figure headings:

Figure 1. The network was trained to recognize objects (as depicted in the top row), and neurons in the hidden layer of the network learned to respond preferentially to image patterns that look like the objects (as depicted in the bottom row). That is, these neurons learned to represent information much like grandmother cells. Taken from [8]

Figure 2. Stimuli used across multiple experiments: a.[31]; b.[21]; c.[35]; d.[36]. e.[46]; f.[47]; g.[39]

Table 1

Year	Paper	Face/Object	Trained to ID images?	Conclusions regarding neural coding:
2015	Dubois, J., de Berker, A. O., & Tsao, D. Y.	Faces, FOB*	No	No mention of GMC** (propose DC***)
2015	Meyers, E. M., Borzello, M., et al.	Faces, FOB	No	No mention of GMC (propose DC)
2015	Taubert, J., Van Belle, G., et al.	Faces	No	Did not discuss coding
2012	Issa, E. B., & DiCarlo, J. J.	Faces	No	Did not discuss coding
2010	Freiwald, W. A., & Tsao, D. Y.	Faces and FOB	No	No mention of GMC (propose DC)
2008	Meyers, E. M., Freedman, D. J., et al.	Animals (cats or dogs)	No	No mention of GMC (propose DC)
2007	Franco, L., Rolls, E. T., et al.	Faces, objects, scenes	No	Did not discuss coding
2006	Tsao, D. Y., Freiwald, W. A., et al.	Faces, FOB	No	No mention of GMC (propose DC)
2005	Freedman, D. J., Riesenhuber, et al.	Animals (cats v dogs)	No	Did not discuss coding
2001	Keysers, C., Xiao, D. K., et al.	Faces/objects/scenes	No	No mention of GMC (propose DC)
2001	Sheinberg, D. L., & Logothetis, N. K.	Objects	No	Did not discuss coding
2001	Tamura H, Tanaka K	Objects, shapes	No	Did not discuss coding
1998	Booth, M. C., & Rolls, E. T.	Objects	No	Reject GMC (propose DC)
1997	Rolls, E. T., Treves, A., & Tovee, M. J.	Faces	No	Reject GMC (propose DC)
1996	Higuchi, S. I., & Miyashita, Y.	Fractal patterns	No	Did not discuss coding
1995	Rolls, E. T., & Tovee, M. J.	Faces	No	Reject GMC (propose DC)
1992	Young, M. P., & Yamane, S.	Faces	No	Reject GMC (propose DC)
1985	Perrett, D. I., Smith, P. A. J., et al.	Faces	No	Agnostic
1984	Rolls, E. T.	Faces	No	Reject GMC (propose DC)
1984	Desimone, R., et al.	Faces and hands	No	Reject GMC (propose DC)
1982	Perrett, D. I., Rolls, E.T., & Caan, W.	Faces	No	No mention of GMC (propose DC)

*FOB: images of faces, objects, bodyparts; **GMC: grandmother cell coding; ***DC: distributed coding

Figure 1

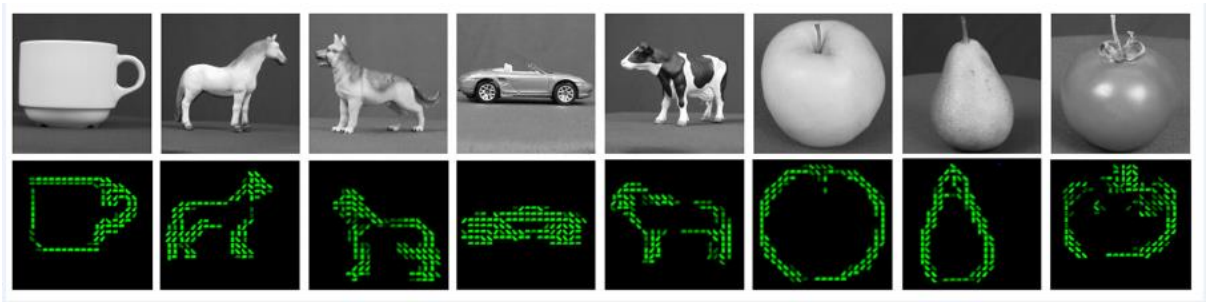

Figure 2